



## UvA-DARE (Digital Academic Repository)

### No data left behind

Visser, I.

**DOI**

[10.1002/icd.2339](https://doi.org/10.1002/icd.2339)

**Publication date**

2022

**Document Version**

Final published version

**Published in**

Infant and Child Development

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Visser, I. (2022). No data left behind. *Infant and Child Development*, 31(5), [e2339]. <https://doi.org/10.1002/icd.2339>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

**COMMENTARY**

# No data left behind

**Ingmar Visser** 

Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

**Correspondence**

Ingmar Visser, Department of Psychology,  
University of Amsterdam, Postbus 15916,  
Amsterdam 1001 NK, The Netherlands.  
Email: [i.visser@uva.nl](mailto:i.visser@uva.nl)

**Handling Editor:** Dr. Moin Syed**Abstract**

Infant research notoriously suffers from small samples, resulting in low power. Beyond increasing sample sizes, improving the reliability of our measurements can also increase power and help find more reliable effects. Byers-Heinlein, Bergmann and Savalei (2021) provide both an analysis of the problem of (low) reliability and a number of valuable recommendations. One of the recommendations is to 'exclude unreliable data'. Although this may increase the effect size found in the remaining data, it can also unjustifiably bias the estimates when it is unknown what the cause of the unreliability is. In such cases, it is better to embrace the variability and use it to characterize the population: variability is also informative. Modern analytical techniques can be used to deal with variability and with missing data. No data should be left behind!

**Highlights:** Variability and individual differences are the bread and butter of developmental science.

Discarding variable/unreliable data carries the risk of biasing effect size estimates.

Variable and missing data can be dealt with appropriately with modern analytical approaches.

Byers-Heinlein et al. (2021) argue that lack of reliability hinders progress in infant research and they provide recommendations for improving reliability. This is indeed much needed: better measurement instruments, more data/trials per infant, and better reporting of the psychometric

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Author. *Infant and Child Development* published by John Wiley & Sons Ltd.

properties of measurement instruments will improve inference in individual studies and in our field as a whole. One recommendation however—‘exclude low quality data from analysis’—is risky and unnecessary.

**KEYWORDS**

cognitive mechanism, individual differences, missing data, mixture models, reliability, variability

## 1 | COGNITIVE MECHANISM AND INDIVIDUAL DIFFERENCES

The goal of much infant research is to establish a particular experimental effect, for example, a difference in looking time between the habituated stimulus and a novel stimulus. In such situations, the largest effect size will be obtained when all infants behave identically, that is when there is no measurement error whatsoever. Note that there is a paradox looming here: when all infants behave identically, apparently there are no individual differences. In experimental psychology studying adults performing in psychophysics tasks, getting rid of individual differences as much as possible may be an acceptable strategy, in infant research it is not. It is throwing away the proverbial baby with the bathwater. Without individual differences, there can be no study of development. Infants acquire skills in different orders, at different points in time and older infants respond faster and more reliably than younger infants. Aiming for a situation where those differences do not exist is hence not the way forward, variability and detecting effects or cognitive mechanisms underlying them need to be reconciled; see, for example, discussion in Borsboom (2006) and an example in decision making in van der Maas et al. (2011).

For such reconciliation to be achievable, good quality measurement instruments are certainly needed! The use of eye-tracking holds a great promise in this respect, for a number of reasons. Data from eye-tracking are much richer than looking times; the whole pattern of fixations and saccades over time can be analysed (e.g., van Renswoude et al., 2020) providing a basis for analysing strategy differences (Kucharský et al., 2020). Basic eye-tracking measures are known to have good test-retest reliabilities (Wass et al., 2014). Moreover, eye-tracking tasks can be made to be more engaging for infants by making them gaze-contingent (Wang et al., 2012), and as such also naturally provide more datapoints per infant. All of these changes can increase the chances of finding reliable and (relatively) large results.

## 2 | DEALING WITH NOISY AND MISSING DATA

Infant data is notoriously noisy, and even with improved measurement procedures part of this variability will remain, as well as the fact that some infants will complete more trials than others. If the goal is to increase effect sizes per se, one could argue that leaving out such unreliable data will have the desired effect. However, this carries a serious threat to the validity of the research as well. In the example that Byers-Heinlein et al. (2021) discuss from the ManyBabies 1 study, it is shown that the effect size increases when only infants are analysed who completed more trials. It is not known however (or at least it is not part of the analysis) what the reason is that some infants complete fewer trials than other infants. Did they just happen to be hungry or thirsty and therefore completed fewer trials? That is, was the reason that they completed fewer trials something outside of the task they were performing or not? Alternatively, it could have been the case that they were less attentive because of some cognitive processing reasons. The infants could potentially already have habituated to the test trials sooner than other infants. Or, yet another option is that these infants were less able to concentrate on these particular stimuli. Yet another possibility is simply that

they were slightly younger than other infants who completed more trials; younger infants are typically more variable, and may therefore complete fewer trials. The truth of the matter is that in many cases as infant researchers we will not know what the reasons are that some infants complete more trials than others. In such cases, it is not advisable to discard data from infants with fewer trials even though that would increase the effect size.

The effect of discarding data is potentially that the effect size estimate becomes biased. For example, if indeed the number of completed trials is dependent on age—which is quite likely in actual experiments!—discarding data based on this will result in an effect size estimate that is biased towards older infants. It should also be noted here that large effect sizes are not a goal in and of itself. In research, we need to establish an effect to draw conclusions about the existence of theoretical constructs and their relationships; there are many cases where we expect small effects (even when our measurement instruments are optimized)—when these putative effects are theoretically interesting enough, we simply need to accept that we will require large samples to detect them. Rather than discarding unreliable data, it is more informative in this case to apply the sixth recommendation in Byers-Heinlein et al. (2021): the employment of more sophisticated analytical techniques. A number of further recommendations can be made about ways in which more value can be gained from variable, developmental data.

First, even prior to doing any analysis, it is advisable to *refrain* from averaging trial data from infants. By averaging trial-based data per infant, differences in variability and reliability are obscured, where we would rather analyse them, for example, in random effects models. If significant random effects are found, follow-up analyses can look into the causes of such variability, for example, related to age, trial number, testing condition, and/or the strength of attention in the pre-test (habituation, familiarization) phase of the task.

Second, consider using latent group or mixture models to capture heterogeneity. In random effects models, participant related variability in the data is thought to result from a continuous, typically Gaussian, distribution. In mixture models, individual differences are conceptualized as qualitative differences, rather than continuous differences. Schaaf et al. (2019) apply a hierarchical Bayesian approach in a reinforcement learning task with mixture components that represent (1) participants that learn in the task and (2) participants that fail to learn and remain at guessing level throughout the trials. Note that such patterns of data may also be observed in habituation, where one group habituates and another group does not. Mixture analysis can be used to detect groups of participants that behave similarly without making prior assumptions about the characteristics of these groups (Van Der Maas & Straatemeier, 2008).

Third, even missing data can be informative! Beyond dealing with missing data appropriately in applying analytical methods (cf Nicholson et al., 2017), the pattern of missing data can be informative, for example when it correlates with age—which is quite likely in developmental data—but also when it correlates with task condition. It could be that task difficulty plays a role, leading to more missing data in one condition than another, which is informative about the nature of the task. Analysing the number of completed (or missing) trials, possibly in interaction with age or other variables, can hence be informative about the task. The pattern mixture model has been developed for analysing such patterns of missing data (Hedeker & Gibbons, 1997).

Discarding data is unnecessary and risky, no data should be left behind! Even when data are variable (or missing!) it can carry information about the cognitive processes that we are interested in. Getting familiar with the more sophisticated analytical techniques required to deal with variable data can be a daunting perspective. Fortunately, our field is moving towards larger collectives of researchers collaborating (Byers-Heinlein et al., 2020; Frank et al., 2017). In such collectives, a division of labour can be more easily arranged and each member can focus on their specialties.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/icd.2339>.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

Ingmar Visser  <https://orcid.org/0000-0003-3855-2778>

## REFERENCES

- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425–440.
- Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., Kominsky, J. F., Kosie, J. E., Lew-Williams, C., Liu, L., Mastroberardino, M., Singh, L., CPG, W., Zettersten, M., & Soderstrom, M. (2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology/Psychologie Canadienne*, 61(4), 349–363.
- Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). Six solutions for more reliable infant research. *Infant and Child Development*, 31(5), e2296. <https://doi.org/10.1002/icd.2296>
- Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Flocchia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., & Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, 22(4), 421–435.
- Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2(1), 64–78.
- Kucharský, Š., Visser, I., Trușescu, G.-O., Laurence, P. G., Zaharieva, M., & Raijmakers, M. E. (2020). Cognitive strategies revealed by clustering eye movement transitions. *Journal of Eye Movement Research*, 13(1-1), 1-20.
- Nicholson, J. S., Deboeck, P. R., & Howard, W. (2017). Attrition in developmental psychology: A review of modern missing data reporting and practices. *International Journal of Behavioral Development*, 41(1), 143–153.
- Schaaf, J. V., Jepma, M., Visser, I., & Huizenga, H. M. (2019). A hierarchical bayesian approach to assess learning and guessing strategies in reinforcement learning. *Journal of Mathematical Psychology*, 93, 102276.
- van der Maas, H. L., Molenaar, D., Maris, G., Kievit, R. A., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review*, 118(2), 339–356.
- Van Der Maas, H. L., & Straatemeier, M. (2008). How to detect cognitive strategies: Commentary on “differentiation and integration: Guiding principles for analyzing cognitive change”. *Developmental Science*, 11(4), 449.
- van Renswoude, D. R., Raijmakers, M. E., & Visser, I. (2020). Looking (for) patterns: Similarities and differences between infant and adult free scene-viewing patterns. *Journal of Eye Movement Research*, 13(1-2), 1-20.
- Wang, Q., Bolhuis, J., Rothkopf, C. A., Kolling, T., Knopf, M., & Triesch, J. (2012). Infants in control: Rapid anticipation of action outcomes in a gaze-contingent paradigm. *PLoS One*, 7(2), e30884.
- Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy*, 19(5), 427–460.

**How to cite this article:** Visser, I. (2022). No data left behind. *Infant and Child Development*, 31(5), e2339. <https://doi.org/10.1002/icd.2339>